DOCUMENT RESUME

ED 026 083 LI 001 132

By-Katter, Robert V.
Transmission Fidelity in Document Systems.
System Development Corp., Santa Monica, Calif.
Report No-SP-3231
Pub Date 1 Oct 68
Note-12p.; Paper presented at the American Psychological Association Convention (San Francisco, September, 1968).
EDRS Price MF-$0.25 HC-$0.70
Descriptors-*Abstracting, *Automation, Cataloging, *Classification, Communication (Thought Transfer), Content Analysis, *Evaluation Methods, Experimental Psychology, *Indexing, Information Processing, Information Retrieval, Information Systems, Libraries
Identifiers-*Pattern Congruity Method

Document searching systems, such as those in libraries, are comprised of chains of components that transmit information. The components consist of mechanisms for doing classification, indexing, abstracting and cataloging of documents, and for converting users' statements of need to expressions suitable for searching the system. Improvement of the transmission fidelity of components requires that fidelity be measurable. The Pattern Congruity Method has been devised for accomplishing this. The method adapts paired-comparison judgment procedures from experimental psychology to the measuring of judgments of document content and the content of representations such as titles and abstracts. It provides a measure of agreement between different instances of the same component, such as between indexers, and also allows direct measurement of fidelity, i.e., the degree of correspondence or congruity between patterns of content of full documents and patterns as depicted by each component in the chain that transduces information about document contents. The method has been used to compare human and machine indexing products, to evaluate machine classification, to study conceptual structures underlying the use of indexing terms, and is currently being employed to study functional overlap in abstracting and indexing services. The method has been experimentally tested for two years under sponsorship of the National Science Foundation. (Author/JB)

ED 026085

LF 001132

SDC

SP-3231

TRANSMISSION FIDELITY IN DOCUMENT SYSTEMS

Robert V. Katter

1 October 1968

LF 001132

SP-3231

SP *a professional paper*

TRANSMISSION FIDELITY IN DOCUMENT SYSTEMS

Robert V. Katter

1 October 1968

SYSTEM DEVELOPMENT CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA
90406

Presented at the American Psychological Association Convention, September 9, 1968

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

(Page 2 Blank)

ABSTRACT

Document searching systems, such as those in libraries, are comprised of chains of components that transmit information. The components consist of mechanisms for doing classification, indexing, abstracting and cataloging of documents, and for converting users' statements of need to expressions suitable for searching the system. A system can be no more effective than the weakest link in its chain of such components, but to improve the transmission fidelity of components requires that fidelity be measurable. A method has been devised for accomplishing this--called the Pattern Congruity Method, it has been experimentally tested for two years under sponsorship of the National Science Foundation.

The method adapts paired-comparison judgment procedures from experimental psychology to the measuring of judgments of document content and the content of representations such as titles and abstracts. It provides a measure of agreement between different instances of the same component, such as between indexers, and also allows direct measurement of fidelity, i.e., the degree of correspondence or congruity between patterns of content of full documents and patterns as depicted by each component in the chain that transduces information about document contents.

The method has been used to compare human and machine indexing products, to evaluate machine classification, to study conceptual structures underlying the use of indexing terms, and is currently being employed to study functional overlap in abstracting and indexing services.

TRANSMISSION FIDELITY IN DOCUMENT SYSTEMS

About fifteen years ago people began to have happy visions of using digital computers to solve all the problems of document storage and retrieval systems. Gauged against these grand early visions, the results so far have been somewhat disappointing. Many early optimists have been forced into an awareness of the complexity, subtlety, and noisiness of the processes of communicating through the printed word. As might be predicted, the disappointments have changed some ardent early optimists into equally ardent pessimists who now use the words "impossible" and "fruitless" with abandon.

The attitude I would like to share with you this morning is somewhere between the extremes of optimism and pessimism. It consists of changing and scaling down expectations about what can be achieved soon in computer applications to document retrieval systems, and of changing and scaling up expectations about the kinds, amounts, and intensities of empirical observations of human communication behavior necessary to gain the more limited objectives. I believe very successful applications can be developed in the coming decade, but it will not be as cheap or easy as many have supposed.

For a computer to act as intermediary between producers and users of documents, it must employ a scheme for communicating between producers and users about the contents of documents. Indexing and classification schemes are examples. For such a scheme to be effective, it must be thoroughly grounded in current empirical data about the conceptual contents and communicative behaviors that are used in common, or that can be made to be used in common, by the communicating parties. Several questions that are unanswered at present are; how much of what kind of empirical data need to be gathered, how often, and in what form?

Psychologists might naturally think of investigating psychological judgment methods and models for answering these questions, and that's what we at System Development Corporation are doing. From the psychological viewpoint such studies can stimulate new theoretical and applied problems, and generate justification and support for trying to solve them. From the point of view of documentation science and technology, such studies can provide new, perhaps better-organized, conceptual models for document-communication processes, better means for relating such models to relevant events, and better means of studying the relevant events.

We have been studying judgmental responses in relation to hierarchical classification models, network clustering models, and multi-dimensional Euclidean classification models. The two studies I wish to describe here used the method of multi-dimensional scaling of paired-comparison judgments.

One study compared the fidelities with which titles, index term assignments, abstracts, and digests portray the main contents of journal articles. The other study was a demonstration of how coordinate indexing schemes contain much presently untapped power to communicate. Both studies were supported by the National Science Foundation and System Development Corporation.

In the paired-comparison procedure a set of perhaps a dozen stimulus objects is selected, such as journal articles, abstracts of those articles, or titles of those articles. The judge is presented with all possible pairs of objects in the set, a pair at a time, and rates each pair on a nine-point scale for the degree of perceived content similarity between the stimulus objects. The numerical rating values so obtained can be thought of as expressing either degree of similarity or degree of difference. These values are subjected to multi-dimensional scaling analysis, which determines how many independent Euclidean dimensions of judged content are required to summarize the rating data, and arranges each stimulus object for its value or position on each dimension. In essence, the procedure provides a summary picture of the main dimensions of judgment along which the judge has distinguished between the stimulus objects.

In the first study, we used an adaptation of the above procedure to measure the content-transmission fidelity of titles, abstracts, index descriptions and digests for journal articles. The sense in which we are using the phrase "transmission fidelity" is very much like that in electrical amplification systems. This meaning is illustrated in Figure 1, which portrays the results of an input-output test of an electrical amplification circuit that shows perfect fidelity. The vertical axis of the input and output graphs represent measured current flow. (T) is an arbitrary reference point in time and (x) is the amount of time it takes for a signal to go through the circuit. The value of the current that is input at time T is amplified to the value that is output at time T plus x, and so on.

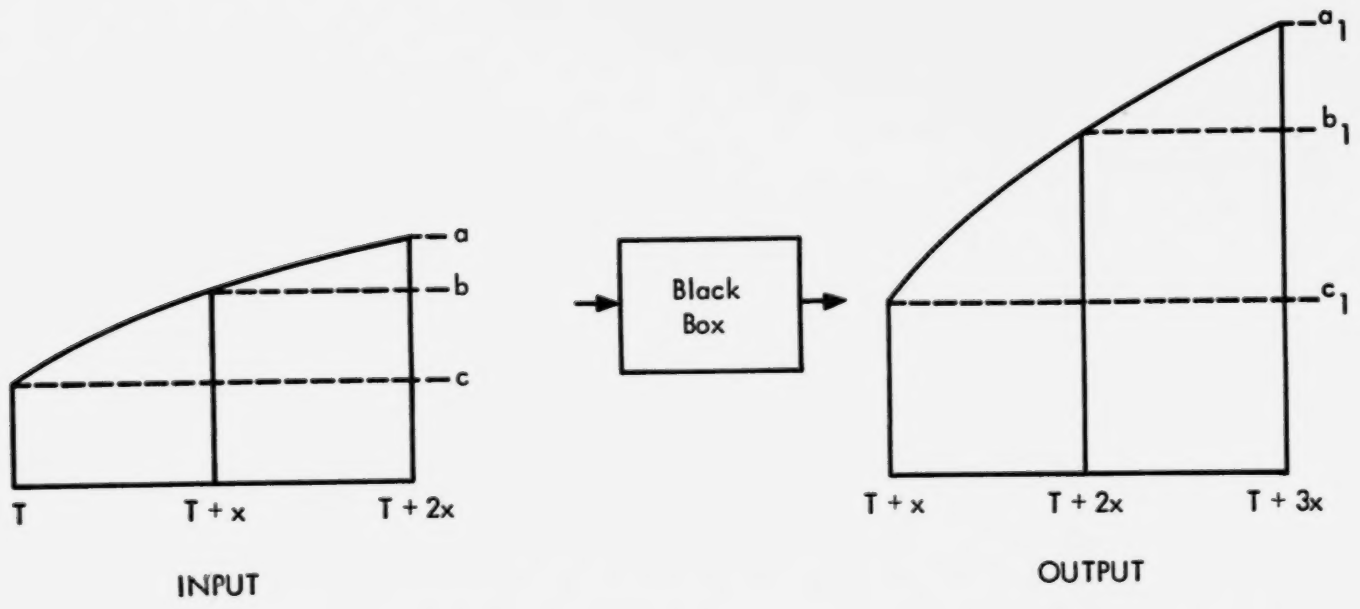

FIGURE 1. SCHEMA FOR TEST OF AMPLIFICATION FIDELITY

On the input graph we can measure the differences a-b, a-c, and b-c, and express the ratios a - b: a - c, a - b: b - c, and a - c: b - c. We can do the same for the differences on the output side. For the amplification circuit to show perfect fidelity, the three corresponding ratios must be identical on the input and output sides of the circuit. The measure of amplification fidelity is, in fact, a measure of the degree of congruity between patterns of corresponding differences between signals measured on the input and output sides of the circuit.

The same kind of strategy can be used to measure the fidelity with which any procedure for constructing representations of documents--such as a particular indexing scheme, a classification scheme, or a standard abstracting practice--transmits document content.

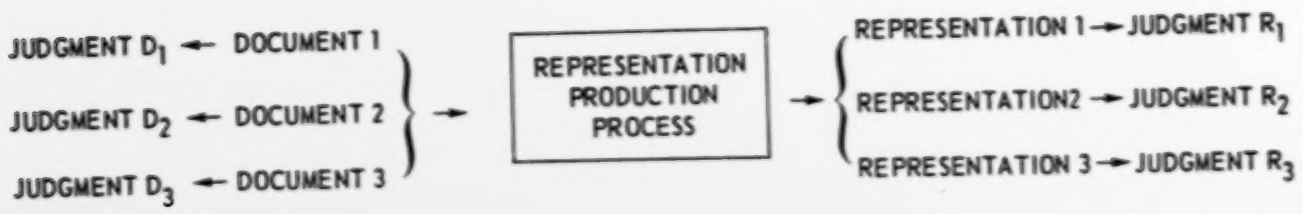
JUDGMENT D_1 ← DOCUMENT 1
JUDGMENT D_2 ← DOCUMENT 2
JUDGMENT D_3 ← DOCUMENT 3
REPRESENTATION PRODUCTION PROCESS
REPRESENTATION 1 → JUDGMENT R_1
REPRESENTATION2 → JUDGMENT R_2
REPRESENTATION 3 → JUDGMENT R_3

FIGURE 2. SCHEMA FOR TEST OF REPRESENTATION FIDELITY

In Figure 2, the black box being tested for fidelity is such a procedure, which, in this case, is a component of a document retrieval system. This component consists of the procedural rules and personnel by which a certain type of document representation is produced. In this example three documents have been input to the process, and three corresponding representations have been output. The main difference between testing the black box in this figure and in the previous one is that the measures taken on the input and output "signals" of this black box are judgmental rather than electrical. Instead of differences in current flow, we have differences in judged degree of content similarity. The judgments of differences are obtained by the method of paired comparisons, which is applied to the full text documents submitted to the input side of the production process, as well as to the corresponding set of document representations that are output. The higher the degree of congruity between the patterns of differences on the input and output side, the more fidelity in transmitting document content.

In the case of the electrical amplification circuit, the congruity of input and output is being measured on a single dimension of current flow. In the case of testing document representations, the congruity between the raw patterns of judgments for input and output is being measured with respect to a space of unknown dimension. However, when the input and output patterns are each subjected to multi-dimensional scaling analysis, the measure of congruity can also be taken on judgmental dimensions.

In our study we made comparisons of the fidelities of digests, abstracts, index-term descriptions, and titles for journal articles in psychology. The results we obtained make intuitive sense; Titles did a pretty good job of representing or transmitting the strongest judgmental content dimension of the full-text articles, but the power of titles to represent the second and third strongest content dimensions decreased sharply. Digests, on the other hand, did the best job of representing all three of the strongest dimensions, with the strongest being represented with the most fidelity, the next strongest with the next most fidelity, and so on. Abstracts showed a little less overall fidelity than digests, and showed markedly decreased fidelity in representing the third strongest dimension. Index term assignments showed approximately equal fidelity in representing the first, second, and third strongest dimensions, but their overall level of fidelity was not so high.

As I mentioned earlier, this pattern of results squares very well with intuitive ideas about what it is that titles, index term descriptions, abstracts, and digests ought to be able to do, and the results thus provide encouragement to look for various points in document searching systems where pattern-congruity testing procedures might be applied.

Figure 3 provides an indication of what these points might be. This is a generalized sketch of a document searching system. The items labelled 1, 2, 3, 4, 5, and 6 are directly observable objects and processes of the system. Item 3 is the input to the process we were just examining in the previous figure, and item 4 is the output. Items 1 and 2 are the input and output for the request-processing part of the system. Item 5 is a directly observable mechanical matching between indexing term descriptions of requests and indexing descriptions of stored documents, and item 6 is the list of document citations that the system finally delivers.

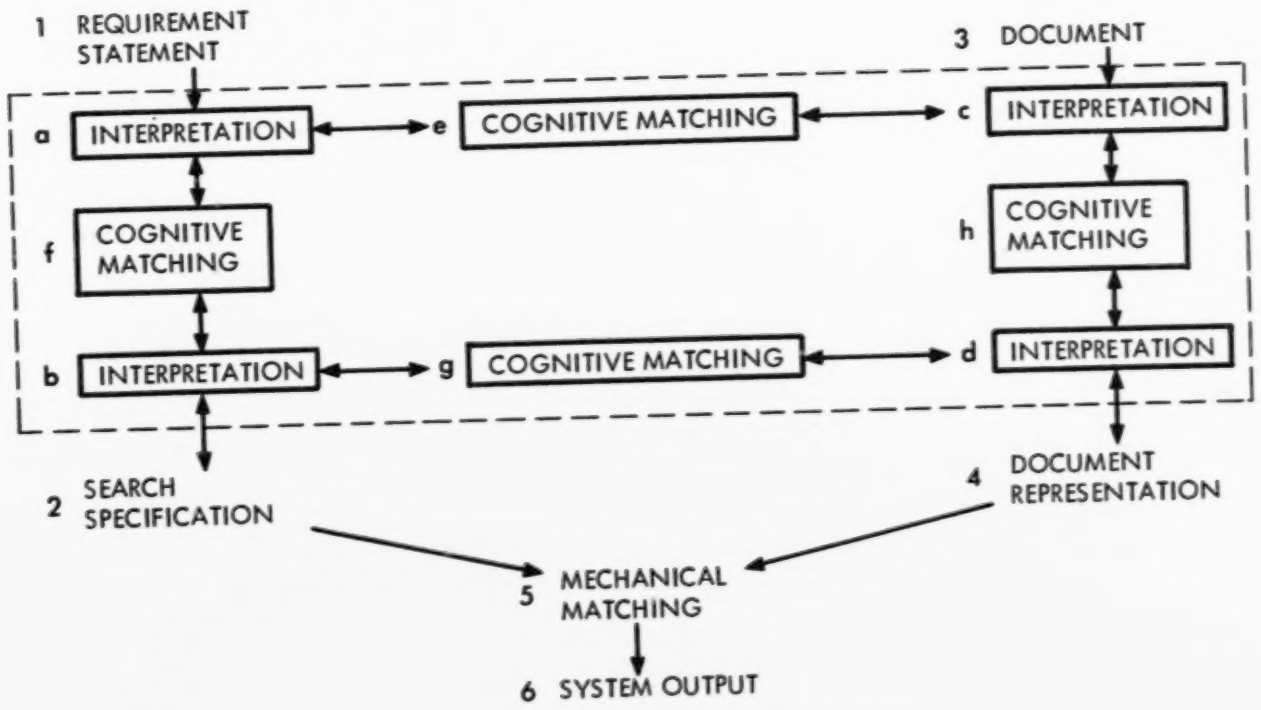

FIGURE 3. DOCUMENT SEARCHING CHAIN

Notice that the process of producing a document representation, which we previously diagramed as a single-stage process, is now expanded into a chain of three elements--c, h, and d--inside the dashed-line box. Each element consists of a judgmental or interpretive process, the nature of which must be inferred on the basis of behavioral measures such as the paired-comparison judgment procedure. Another point can now be made clear; the upper limit of possible fidelity of any process that produces document representations will be a product of the fidelities of the three elements c, h, and d, (i.e. cx hxd). Since this is a social communication system, a good measure of the fidelity of such elements is the average congruity between appropriate judges, both with respect to document judgments and document-representation judgments. We investigated such between-judges interpretive reliabilities, and as you might guess, found them highest for the essentially uni-dimensional titles, with the descending order of reliabilities continuing down, through index descriptions, abstracts, and digests, to the most content-complex stimulus of them all, the full-text article. In other words, digests, which provide the best interpretive coverage of document contents, are plagued by lower interpretive reliabilities, while titles, which tend to provide coverage of only the more obvious of document contents, show the highest between-judge reliabilities. It is clear that individual-difference factors of selective attention and emphasis are operating more freely in the responses made to full texts and digests than in responses made to titles. There is a challenge to find practicable means of reducing such selective effects in the production of index descriptions, abstracts, and digests.

Equally important for the overall fidelity of the system is to apply the same procedures to the other side of the system, which converts requests for information into search formulations. Finally, the four elements labelled "cognitive matching" are included in the diagram to emphasize that there is no necessarily guaranteed connection between the interpretations in any of the four elements a, b, c, or d. Interpretive reliabilities might, for example, be high in both elements c and d, but the match between the two systems of interpretation could still be quite poor. In considering all these boxed elements, if it gives you the impression that fidelity in large-scale document searching systems may require testing and improvement at many points, then I have gotten my message across.

But what about element 5, the mechanical matching process? We have time for a brief look at another experimental application of multi-dimensional scaling analysis to the problem of whether mechanical matching takes full advantage of all the information available in combinations of indexing terms. Human and machine matching of index terms are not the same. In typical machine applications of coordinate indexing, the mechanical matching procedure matches terms to terms. For example, if the search formulation calls for term A, the system will deliver citations containing term A. If it calls for terms A and B, only citations containing both A and B will be delivered, and if the formulation calls for terms A or B, the system delivers items containing either A or B or both. In a few systems, close synonyms for terms A and B are automatically taken from a thesaurus and added to the search formulation.

Suppose, however, that instead of having a card sorter or a computer to do the matching, we have a person. This person is very well informed in the area of technical content being searched, and also has lots of time and patience. We provide our human matcher with the index term description of the user's request, and with index descriptions of the documents available in storage. He can duplicate the term-to-term matching of the card sorter or computer, but he is also likely to add a few tricks of his own. He will react to the combination of indexing terms in the search formulation as a gestalt, in which terms with several possible meanings are interpreted correctly because of the context of other terms. Also, he will assign more weight to some terms than to others, on the basis of the hypothesis he has developed about what is being expressed by particular combinations of terms. If asked to do so, he will usually be willing to select other terms that are not synonyms for the ones presented but that he thinks would also describe the user's need. He treats the index-term descriptions of available documents in the same way. In short, he builds a concept from the terms of the search formulation, and a concept from the terms describing each document, and he matches the search concept against document concepts.

Is there even a possibility of modelling such activity on a computer? From a behavioral point of view, our well-informed matcher takes an incomplete indexing description of a document or information request, and on the basis of his expectations about probable co-occurences of kinds of information, he makes the description more complete. If his expectations accord with reality, search accuracies are improved. But we may ask, is the well-informed matcher highly dependent on each unique stimulus configuration, or can he reconstruct stimulus configurations when their parts are presented to him in an unorganized form? It is doubtful that a computer program could accomplish useful reconstructions if a well-informed matcher cannot.

To get a first estimate of whether well-informed matchers can perform useful levels of such reconstruction, we studied the behavior of requesters and indexing personnel in two large aerospace information dissemination systems. The following paradigm was essentially replicated twenty-five times, using various controls. First, from among articles selected by the system in response to a particular request for information, one article was identified that the requestor had rated as of high interest, and one as of low interest. Second, the indexing terms describing the high-interest article, the low-interest article, and the user's request for information were all put together into a combined list, and an exhaustive paired-comparison inventory was constructed from this list. Every term on the list was paired once with every other term. Third, a worker from the library who was familiar with the indexing scheme was asked to make judgments of the degree of similarity of document content implied by each pair of terms in the inventory. These judgmental data were subjected to multi-dimensional scaling analysis. None of the judges was aware of the purpose of the experiment, or of how the terms in the inventories had been selected.

Figure 4 illustrates three consistent trends found in the results. In the figure, the positions of the indexing terms on the strongest dimension are plotted on the horizontal axis, with positions on the second-strongest dimension being on the vertical axis. Empty triangles represent terms describing only the low-interest article, empty squares represent terms describing only the high-interest article, and dots represent terms from only the search formulation. The other symbols represent terms that described both articles, or an article and the search formulation.

The first trend illustrated is clustering. First, the bulk of the empty triangles are clustered in the positive sector of dimension I, and in the neutral sector of dimension II. Also, the bulk of the empty squares are positioned slightly negative on dimension I and positive on dimension II. Finally, the dots show two definite clusters on opposite ends of dimension II. The clusterings suggests that the judge did perceive certain combinations of terms as being more likely to co-occur in describing information contents, and that his perceptions corresponded with the realities of the index descriptions of the journal articles and the request.

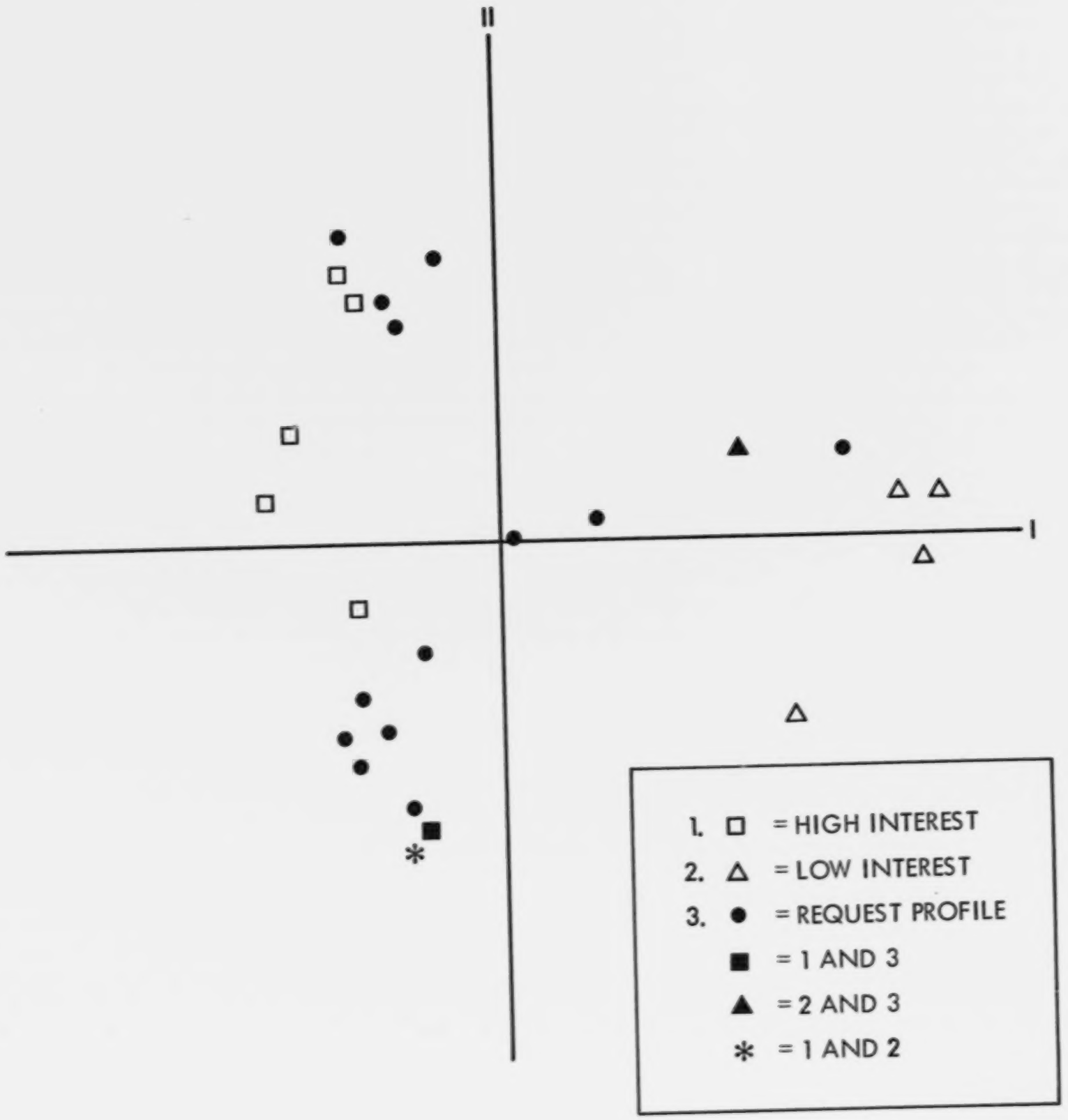

FIGURE 4. MULTIDIMENSIONAL SCATTERPLOT

The second trend is separation: the main clusters from the high and low interest articles are oppositely loaded on the strongest dimension.

The third trend is co-location: the bulk of both the request profile dots and the high-interest article squares are located in the same general position on dimension I.

It seems clear that if a computer matching system contained these blind-judgment relations modelled in an available form, it could easily select the high-interest article and reject the low interest one. Recall that the straight term-matching system was not able to make a distinction between the high and low interest articles. These results seem encouraging to us.

In conclusion, while studies of the kind I have described have not provided full solutions to problems of document representation, we do feel that they help to lay open the problems, and to indicate directions in which more progress is likely to be made.

REFERENCES

1. Katter, R. V. Study of document representations; multidimensional scaling of indexing terms. Final Report. 31 August 1967. TM-3090

2. Katter, R.V., Principal Investigator. Experimental investigations of a method for analyzing document representations. Progress Report. 9 August 1966. TM-3090.